

Habitability and Performance Issues for Long Duration Space Flights

Mihriban Whitmore and Meredith L. McQuilkin
Lockheed Martin Space Mission Systems & Services
Houston, Texas

and

Barbara J. Woolford
NASA Johnson Space Center
Houston, Texas

Acronyms

AFB	Air Force Base
AGARD STRES	Aerospace Research and Development Standardized Tests for Research with Environmental Stressors
CSERIAC	Crew System Ergonomics Information Analysis Center
DSSQ	Dundee Stress State Questionnaire
ESA	European Space Agency
EVA	Extra-Vehicular Activity
FAA	Federal Aviation Administration
FCSD	Flight Crew Support Division
FRANCIE	Framework Assessing Notorious Contributing Influences for Error
HOS	Human Operator System
INEEL	Idaho National Engineering Laboratory
IPT	Integrated Product Team
ISS	International Space Station
IVA	Intra-Vehicular Activity
LDSF	Long Duration Space Flight
LMLSTP	Lunar-Mars Life Support Test Project
LMSMSS	Lockheed-Martin Space Mission Systems and Services Company
MINUTES	MINesota Universal Task Evaluation System

NASA	National Aeronautics and Space Administration
PATS	Psychophysical Assessment Test System
PAWS	Performance Assessment Work Station
POMS	Profile of Mood States
RSA	Russian Space Agency
SA	Situational Awareness
SMEAT	Skylab Medical Equipment Altitude Test
SOIRT	Space Operations Issues Reporting Tool
SWAT	Subjective Work Assessment Technique
THEA	Tool for Human Error Analysis
TLX	Task Load Index
UTAF	Usability and Testing Analysis Facility
UTC-PAB	Unified Tri-Service Cognitive Performance Assessment Battery
WWW	World Wide Web

Habitability and Performance Issues for Long Duration Space Flights

Abstract

Advancing technology, coupled with the desire to explore space has resulted in increasingly longer manned space missions. Although the Long Duration Space Flights (LDSF) have provided a considerable amount of scientific research on human ability to function in extreme environments, findings indicate long duration missions take a toll on the individual, both physiologically and psychologically. These physiological and psychological issues manifest themselves in performance decrements; and could lead to serious errors endangering the mission, spacecraft and crew. The purpose of this paper is to document existing knowledge of the effects of LDSF on performance, habitability, and workload and to identify and assess potential tools designed to address these decrements as well as propose an implementation plan to address the habitability, performance and workload issues.

Habitability and Performance Issues for Long Duration Space Flights

INTRODUCTION

In recent years, the relationship between the U.S. and Russia with respect to space flights has changed from adversarial to a more co-operative approach. Specifically, the space programs are now jointly focus on answering questions about human physical and psychological adaptation to long duration space flight (LDSF) and microgravity. The Russian Space Agency (RSA) continues its presence in space with the permanently inhabited, orbiting space station Mir. In fact, missions aboard Mir have become increasingly longer in duration. Since 1973, there have been over 20 documented manned space flights lasting 60 days or longer (Guidi & Holland, 1992). The current record of 366 days aboard Mir was set by a Russian cosmonaut on a 1987-1988 mission. To continue to foster cooperation between Russia and the United States, NASA and RSA (along with other international partners) have begun work to create a permanently occupied International Space Station (ISS).

The LDSF's conducted to date have provided a considerable amount of scientific research on a human's ability to function in extreme environments. This information has been complemented by studying environments analogous to those experienced during space travel; such as submarines, winter-over expeditions in Antarctica, and other isolated and confined facilities. All of these analogues have helped advance the scientific understanding of human performance in LDSF. Consistent findings from all of these areas of study indicate that long duration missions take a toll on the individual both physiologically and psychologically. Specifically, the isolated and confined conditions aboard space vehicles may induce individual work performance decrements and reduced group socialization skills. These problems may become magnified with increases in mental and/or physical workload. The importance of these findings becomes extremely critical when considering that the minimum individual mission duration aboard the ISS will be three months.

The purpose of this report is to document the existing knowledge of the effects of LDSF on habitability, performance, and workload. In addition, it is intended to identify and assess the potential use of the existing tools for measuring, monitoring the habitability, performance and workload to support identifying the countermeasures in order to minimize the LDSF effects. For the purpose of this paper, habitability is defined as those factors which promote the productivity, well-being, and situationally desirable behavior of crewmembers in space (Habitability Research Group, NASA-Ames' Space Human Factors Office). Performance is defined as the process of

completing a human task in an effective manner. Workload is defined as the amount of physical or mental work assigned for completion during a specified amount of time. The factors affecting habitability may also have an impact on performance and workload. A key factor affecting these three concepts are mission-related such as mission duration, objectives and performance requirements. Other factors that are also as important are crew characteristics such as crew size, selection, cultural background, group characteristics (e.g., size, mix and dynamics), operational elements such as work/rest cycles, external/internal communication, working and living space design, environmental factors such as temperature, noise, reduced gravity, and life support and facilities (e.g., recreational, exercise, medical care) (Salvendy, 1987). In addition, habitability, performance and workload are interconnected and influence each other. Poor living conditions result in a sub-optimal habitat and work environment, which in turn may produce performance decrements. These performance decrements may increase the perceived workload of the crewmember, since more energy must be expended to complete the given task under reduced mental and physical capacity. Likewise, extremely high workload levels can negatively affect performance due to increased demands placed on the limited capacity of the human system. High levels of workload can also influence and exacerbate habitability issues in space. For example, high levels of workload can create stress in crew, which can result in a loss of situational awareness, a decrease in physical comfort, an increased risk of error, and a loss of mission effectiveness. Figure 1 represents a proposed schematic of factors affecting habitability, performance, and workload and how they inter-relate. This proposed schematic is created based on habitability discussions in the Handbook of Human Factors by Salvendy (1987).

Several factors associated with the cognitive capabilities and environment act independently and in concert with one another to impair performance and possibly workload and habitability during LDSF. These factors include, but are not limited to, sensory, neuro/cognitive, psychomotor, decision making/logic/analytical reasoning, emotional state/motivation, psychological adaptation, attention/vigilance/situational awareness, physical adaptation, perceived health, and environmental issues (Holland, 1995). Since the human is the essential element in determining the habitability, performance and/or workload related issues and demonstrating how they all interact and affect, the primary focus of this paper will be to address the issues pertaining to cognitive functioning, attention/vigilance/situational awareness, and factors related to mental workload and performance, and in turn, discuss how they relate to habitability issues during LDSF. Finally, an evaluation and discussion of available performance measurement tools and existing countermeasures will be included as a possible means of identifying and/or reversing performance decrements that occur during LDSF.

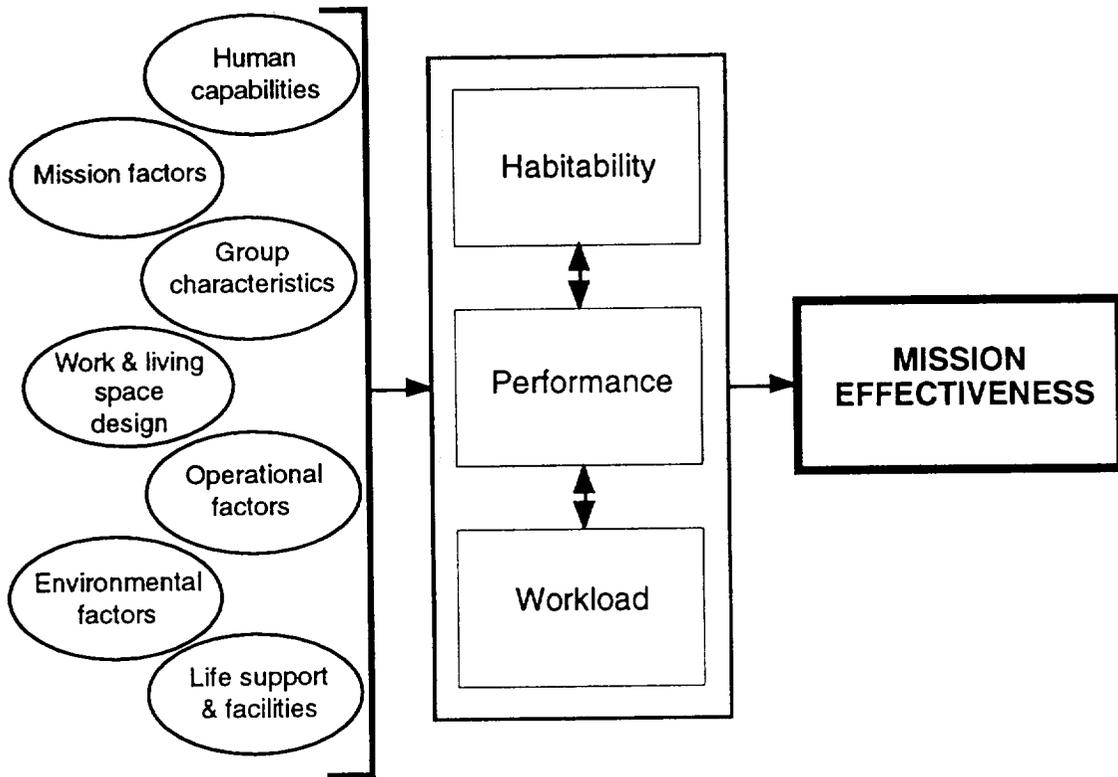


FIGURE 1. Relationship between habitability, performance and workload and factors affecting them.

BACKGROUND

The Crew Station Branch within the Flight Crew Support Division (FCSD) at NASA, has been tasked to prepare an analysis of human performance and habitability issues on LDSF and identify existing tools for use to assess these issues for LDSF missions. The scope of this report was to identify human performance issues during space missions from a crew habitability perspective, and was based on existing literature available at the time of the task. Most of the research focused on performance tools since very little has been written on tools designed to assess habitability issues. The perspective of the discussions within the paper are:

- 1) Was the task completed,
- 2) How efficiently did the human complete the task,
- 3) What were the existing performance - cognitive and perceptual - capabilities and decrements, and
- 4) What impact did relevant habitability issues have on human ability to accomplish the task?

The rationale for this report is multi-faceted: first, ground controllers and mission support personnel should be aware of flight crew performance limitations so assigned tasks can be completed by the crew. Second, an analysis of the current empirical evidence of human capabilities and limitations during LDSF is needed to determine potential relationships between performance factors and tasks. Third, non-intrusive performance and workload measures in addition to accurate and reliable predictors of performance in microgravity environments are needed to assess the impacts of LDSF on performance. Fourth, performance effects from excessive workload in microgravity, as well as solutions to reduce workload, should be assessed to determine countermeasures so that performance degradation can be either eliminated or minimized.

APPROACH

Various resources including, but not limited to, NASA, European Space Agency (ESA), RSA, and Canada were utilized to obtain information about current methodologies used in performance and workload research as well as information on existing human performance and workload measurement tools. A request for such information was posted at various NASA-approved newsgroups and sites on the World Wide Web (WWW). Additionally, an intensive literature search and review was conducted to locate relevant NASA reports (by searching through NASA RECON and NASA GALAXIE databases), technical documents, as well as any research periodicals related to performance and workload. Finally, various researchers, in academia as well as those employed at NASA, were contacted. Additionally, studies and reports from the United States Navy, Federal Aviation Administration (FAA), and Antarctic missions were reviewed for information related to performance and workload. Information gathered has been reviewed and assessed for relevance to this report. Although an attempt was made to gather information from the ESA, RSA, and Canadian Space Agencies, no information from these agencies has been received at the time the paper was written.

Moreover, existing tools to monitor and measure performance and workload were investigated for applicability to LDSF. A preliminary assessment of these tools were conducted. However, these evaluations were limited and more comprehensive evaluations of the candidate tools would be necessary in order to determine their applications for measuring and monitoring crew performance and crew habitability during LDSF. Furthermore, these tools, in conjunction with existing space based procedures, would also need to be evaluated for their use for developing appropriate countermeasures for the detrimental effects of LDSF.

The following sections summarize research findings to date, as well as provide a preliminary assessment of existing tools. Also included are

recommendations for future work on further research measuring issues such as habitability, performance and workload associated with LDSF.

SUMMARY OF RESEARCH FINDINGS

Habitability

Early habitability assessments in microgravity environments were obtained through questionnaires completed by Skylab crewmembers'. Questions ranged from architectural issues to communications and hardware use. The subjective questionnaires and evaluation forms were presented to the Skylab crew in a checklist format through Experiment M487, Habitability/Crew Quarters. This experiment was primarily conducted to evaluate and report on habitability issues found aboard Skylab which would be useful for designers of future spacecraft (NASA TM X-58165, 1975). Data were collected in two specific areas: general aspects of living and working in microgravity, and specific equipment and architectural arrangements of Skylab. One of the major findings of the questionnaire was the importance of habitability provisions for optimal crew performance. For example, Skylab crewmembers spent half of their time on activities, such as personal hygiene and eating, where habitability provisions had a significant effect on the time required to perform those tasks. In many of those cases, slightly more sophisticated equipment or accommodations would have saved crew time (NASA TM X-58165, 1975).

In addition to Experiment M487, Skylab crewmembers also took part in the Skylab Medical Equipment Altitude Test (SMEAT). SMEAT was originally intended to test Skylab flight hardware, crewmembers realized that issues, such as maneuverability and range of motion, would affect performance and habitability, they began using the test to assess habitability concerns and equipment. Unfortunately SMEAT was never used to assess habitability issues following this flight.

The "Space Station Habitability Report" (Boeing, 1983) also discusses the relationship between performance and habitability. Specifically, the report concludes that normal working hours should not impact private time. A Shuttle crewmember was quoted as saying that in flight, "It is frustrating not being able to get everything done. This leads to cutting down on sleep, eating on the run, and not exercising properly, which only add to the frustration". Note that the crewmember recognizes that excessive workload leads to poor performance and decreased morale. Another crewmember who also recognized the relationship between workload and performance indicated that mistakes were often made because the crew was rushed to complete tasks. Furthermore, he stated, "Things that were well understood got botched up because of the mental state we were in". This report suggested that crewmembers be trained to deal with the stresses associated with

extended duration of isolation; however, it did not suggest a specific methodology for the assessment and management of stress during LDSF.

With the advent of American astronauts living onboard Mir for extended periods of time, NASA has developed measurement tools to assess the issues addressed within the Boeing report. Specifically, NASA Space Station Operations and Execution Planners have developed a general questionnaire to be used by MIR crewmembers. This questionnaire is designed to help evaluate crew time planning philosophies for the International Space Station (ISS). NASA is particularly interested in the amount of time needed versus the amount of time delegated for completion of a given task. This tool has shown that if the time required to complete a task exceeds the amount of time allotted, perceived mental workload and mental stress increase while performance decreases. It is unknown if any statistical analysis of the data was performed, or if the findings are based on crew comments and observations.

Another NASA measurement tool developed by the Usability Testing and Analysis Facility (UTAF) to assess LDSF performance and habitability issues is the Space Operations Issues Reporting Tool (SOIRT). SOIRT is used for describing habitability and performance-related incidents during flight. The SOIRT was most recently used during the Lunar-Mars Life Support Test Project (LMLSTP, Phase IIa and III), and two Shuttle missions postflight. Work on documentation of the Phase IIa findings is in progress.

Performance

Various methods and tools have also been developed to assess in-flight performance. Two neuro-cognitive tools, the NASA Performance Assessment Workstation (PAWS) and Department of Defense (DoD) Unified Tri-Service Cognitive-Performance Assessment Battery (UTC-PAB), were developed to evaluate the effects of microgravity on crew cognitive performance abilities. In preparation for the Spacelab Life Sciences (SLS-3) mission, a ground-based study was conducted to assess the impact of sub-optimal training schedules and testing lapses on baselined performance stability (Schlegel, Shehab, Schiflett, Eddy, & Gilliland, 1994). A secondary purpose of the study was to evaluate alternative mission test schedules. Subjects were assigned to various training schedules that involved either 6, 15, or 16 sessions. Following training, subjects experienced a 3 or 5 day lapse between training and testing. Results indicated that there were few performance differences due to the differing training schedules. However, in 9 of the 28 performance measures used, a testing lapse (either 3 or 5-day) resulted in a significant performance decrement (Schlegel et al., 1994).

Another tool, the Psychophysical Assessment Test System (PATs) is a micro-computer based system which provides a comprehensive

measurement of psychophysiological data and can be used in a wide variety of applications ranging from operational environments with 'real-world' tasks to laboratory-based standardized tests. PATS was developed by the Performance Assessment Branch of the Human Engineering Division at Armstrong Laboratory, Wright-Patterson Air Force Base (AFB), and is available through Crew System Ergonomics Information Analysis Center (CSERIAC).

Another area of concern during LDSF involves psychomotor performance. During a series of studies, Manzey, Lorenz, Schiewe, Finell, & Thiele (1993) used a set of four tasks from the battery of Standardized Tests for Research with Environmental Stressors (published by the NATO Advisory Group for Aerospace Research and Development (AGARD STRES)), to test logical reasoning and decision-making, memory retrieval, tracking, and fine manual control. This test battery included a wide variety of cognitive and psychomotor process performance tasks. Pre- and post-testing were conducted on the ground and compared to testing conducted on an 8-day Shuttle flight. Although the results revealed no deficits in short-term memory and logical reasoning, clear decrements were found in tracking performance and fine manual control. The authors believe the performance decrements were a result of "alterations requiring an effortful accommodation of motor skills which had been acquired under 1-g conditions to the new conditions of microgravity" (Manzey et al., 1993). Results also showed that the first and largest performance decrement occurred during early stages of the mission. By the middle of the mission, the crew displayed a clear habituation effect, while performance, declined toward the mission conclusion. According to Manzey et al. (1993), the initial performance decrement occurred because "qualitative changes in the human psychomotor system under microgravity led to an accommodation of the previously learned motor skills" in the new environment. Furthermore, according to Kozlovskaya, Burlachkova, Ganchev, Gatev, Gerstenbrand, & Berger (1993), performance decrements can be explained by severely altered proprioceptive feedback that is often found in 0-g environments. This lack of feedback must be compensated by enhanced visual and/or attentional control of voluntary movements. Consequently, lowered tracking performance observed toward the end of the flight can be attributed to increased mental fatigue. In other words, the increase in mental workload was used as a compensation mechanism for the reduction of proprioceptive feedback. This finding has huge implications for mental workload on LDSF. For instance, the more fine motor control tasks that the crew must perform during a LDSF, the earlier the onset and greater may be the effects of mental fatigue.

Manzey et al. (1995) also used the AGARD STRES to assess the effects of microgravity on dual task performance. One crewmember self-administered a dual-task (unstable tracking with concurrent memory) on 13 occasions during an 8-day mission. Preflight, in-flight, and postflight performance data

were compared. The results of this task demonstrated that over the course of the flight, single task tracking and dual task performance were negatively affected by space flight.

Finally, Manzey et al. (1993) conducted experiments to assess the effects of microgravity on information processing and performance. During an 8-day mission, one crewmember completed a series of tests (Grammatical Reasoning Task, Memory Search Task, and Unstable Tracking Task) from the AGARD STRES battery. Results indicated that logical reasoning functions, as well as the speed and accuracy of short-term memory retrieval, were not affected by the space environment.

Other LDSF factors which impact performance and habitability are mood (emotional state) and motivation. J. Raglin (University of Wisconsin) has extensively studied the effects of over-training on motivation and performance in athletes. Some of his findings can be applied to the intensive training endured by crewmembers before embarking on a LDSF. Specifically, Raglin (1993) and Raglin & Morgan (1994) found that over time intensive physical training, or overtraining, resulted in decreased athletic performance. Additionally, the overtraining produced "significant mood disturbances" (depression, apathy, lack of motivation) which resulted in serious performance decrements. Raglin (1993) recommends rest as a viable solution to the mental and physical problems associated with overtraining. To predict which athletes would succumb to training-induced stresses, Raglin and Morgan (1994) developed a measurement scale for prediction based on items taken from the Profile of Mood States (POMS). The seven POMS items served as predictor variables for the measurement scale developed by Raglin and Morgan (1994).

Another method for determining the level of emotional/ cognitive stress includes speech analysis. By remotely monitoring the vocalizations of flight crews on LDSF, it is possible to use speech measures to determine the workload and stresses exhibited by the crew. To date, it is known that the RSA has used this technique with the Mir cosmonauts, but the results of the analysis are unknown. The interest in using speech as a tool for measuring stress can be attributed to four factors:

- 1) a signal can be obtained from a crew located elsewhere
- 2) the speech signals are acquired in a noninvasive, unobtrusive manner without the need for additional equipment
- 3) speech is produced naturally and frequently by crews
- 4) speech measures can be applied in both real time and after the fact (Stuster, 1996).

Stuster (1996) noted that the most promising speech variables appeared to be speech fundamental frequency, also referred to as pitch, and vocal jitter, which can be seen through cycle-to-cycle variation in the period of the signal. Furthermore, Doherty (1991) reported that although pitch is positively

correlated with stress, there is a weak, negative relationship between vocal jitter and workload. However, it is important to note that neither pitch nor vocal jitter can distinguish between differing levels of stress. Despite the measure's insensitivity to different stress levels, the use of speech patterns as a measurement tool is promising. More research on the effectiveness and reliability of this technique needs to be done before it can be used to evaluate crew stress during LDSF.

On a psychological level, crew adaptation to microgravity could lead to habitability problems and performance decrements. For example, during LDSF, sleep patterns can be severely altered. Just recently, Russian Space Officials attributed the irregular heartbeat of a cosmonaut on Mir to a shift in the crewmember's sleep schedule combined with stresses associated with the accumulation of several traumatic events onboard. This cosmonaut questioned his own ability to perform very difficult and essential tasks for a planned Extra-Vehicular Activity (EVA) repair mission. The Russian Flight Surgeon recommended that the crewman relax by getting plenty of rest and sleep.

Scientific literature has repeatedly documented how sleep decrements often lead to inhibited cognitive functioning and reduced performance. Crewmembers recognize the value of a good night's sleep while in-flight. In one instance, a Skylab crewmember even went as far as to position his sleeping bag closer to the air vent so he could sleep better. Ground studies on sleep requirements and deficits have shown huge implications for LDSF. Existing knowledge of sleep patterns can be applied to space missions to reduce the likelihood of performance decrements due to sleep deprivation.

Blagrove, Alexander, & Horne (1995) studied the effects of chronic sleep reduction on the performance of tasks sensitive to sleep deprivation. They found that subjects, reduced to a mean of 5.2 hours of sleep per night for 28 nights, 4.3 hours per night for four nights, and 5.3 hours per night for 18 nights, showed no logical reasoning or auditory vigilance performance decrements, but did show decrements on their ability to ignore distracting information. Haslam (1982) found that over three days of total sleep deprivation, riflery skills of infantry soldiers showed little deterioration while cognitive performance accuracy dropped each day. Furthermore, Elsmore, Hegge, Naitoh, Kelly, Schlangen, & Gomez (1995) found that when sleep debt accumulates, task performance continues to follow a circadian rhythm while degrading linearly, especially on tasks low in motivating qualities. In addition, Haslam (1982) found that performance during the circadian trough may be slower to return to baseline after recovery sleep periods.

Due to the importance of proper sleep patterns on performance, several researchers have studied the effectiveness of various countermeasures on sleep deficiencies. For example, Mitler, Carskadon, Czeisler, Dement, Dinges,

& Graeber (1989) found that physical activity and dietary stimulants can temporarily mask fatigue and Bonnet & Arand (1994) showed 200 mg of caffeine benefits response times, vigilance, and alertness during periods of sleep deprivation. While not long term solutions to sleep decrements, these measures can provide temporary solutions during critical periods when the crewmembers must be alert and functioning near capacity.

Recently, much attention has been given to the situational awareness of crews during LDSF. Holland (1995) writes that during LDSF isolation and confinement stressors "tend to enhance situational outcomes that might not otherwise occur. It is these outcomes which have the potential to degrade human performance and interactions" among crewmembers. He has defined 5 levels of group situational awareness (SA):

- 1) normal group SA
- 2) slightly-impaired SA
- 3) moderately-impaired SA
- 4) severely-impaired SA
- 5) loss of group SA

Finally, he states that "knowing when the functional sub-group situation awareness needs to be high and when the total team awareness must be high is very contextually dependent and should be considered thoughtfully as an operational mission parameter and research issue". However, despite the importance of situational awareness during LDSF, little documented research, including tools of assessment, has been conducted to investigate the causes and effects of a lack of situational awareness. There needs to be more research into the feasibility of developing and using assessment tools of situational awareness.

Workload

Perceived mental workload is probably the major psychological factor affecting crew performance during LDSF. According to Manzey (1989) mental workload should "not be considered as a single quantity, but as a multidimensional concept (there are as many workloads as resources)". Additionally, workload assessment techniques "should be able to clarify the demands (workload) of a certain task with regard to each of the different kinds of processing resources" (Manzey, 1989). One of the most common methods of assessing mental workload is through the secondary task technique. This technique requires the simultaneous performance of two tasks: the primary task to be evaluated with regard to the demands placed on the operator; and a secondary task, the performance of which is usually taken as an indicator of "spare capacity" (Ogden, Levine, & Eisner, 1979). It is important to note that the secondary task mainly taps a certain kind of processing resource, (cognitive or response-related), and that the task is not data limited, meaning the task performance varies depending on the availability of the 'spare' resources. Theoretically, the greater the mental

workload needed for completion of the primary task, the worse the performance on the secondary task. However, these results might differ depending on the amount of resource overlap between the primary and secondary tasks. Specifically, if the primary and secondary tasks both use problem-solving resources, the secondary task might suffer in performance compared to a secondary task that requires resources devoted to less cognitively driven resources, such as stimulus-response actions.

In ground-based studies, Hancock and Williams (1993) used the MINesota Universal Task Evaluation System (MINUTES) and the Subjective Work Assessment Technique (SWAT) to assess the effects of task load and task load increments on performance behavior. Results show a rise in task load produces an increase in both the time to react correctly to a monitoring cue and the number of false responses, both of which indicate a deterioration of capability. Unfortunately, these tools requires complete concentration on the task by the subject and at this point is not feasible for use by crewmembers while in-flight.

Hart and Staveland (1988) created the NASA-TLX (Task Load Index) which is a subjective workload assessment tool designed to assess performance while completing various tasks. The workload measurement is dependent on a weighted average of six factors found to contribute to workload:

- 1) mental demand (the level of mental and perceptual activity required to complete a task)
- 2) physical demand (the level of physical activity required)
- 3) temporal demand (the level of time pressure)
- 4) performance (satisfaction about own performance)
- 5) effort (how hard someone has to work)
- 6) frustration level (discouragement, irritation, annoyance, etc.).

Veltman & Gaillard (1996) used the NASA-TLX to assess physiological workload reactions to increased levels of task difficulty and found that heart period, heart rate variability, blood pressure, and respiration all reflect large differences in mental effort.

The Dundee Stress State Questionnaire (DSSQ) is a measurement tool recently developed by researchers in Dundee Scotland, to assess workload. DSSQ findings indicated workload was mostly derived from mental demands and least from physical demands. Furthermore, workload was found to be closely related to tension and least to motivation. Pleasantness of mood, self-focus, perceived control, and self-esteem did not seem to have as great of an impact on workload as predicted. The authors posed two hypotheses to explain their findings:

- 1) the person's experience of the workload imposed by the task may have affected subjective state

- 2) the person's subjective state may have biased appraisal of task demands and personal reaction to the task (Matthews, Campbell, Joyner, Huggins, and Falconer, 1993).

A tool used for predicting mental workload is a computer-based software program called MicroSaint that was developed by Micro Analysis and Design. MicroSaint, based on a dynamic human performance model, decomposes a task to find instances of behavior across time where task demands are high. Additionally, the same company has incorporated MicroSaint with another tool, the Human Operator Simulator (HOS), to create a way to bridge the gap between anthropometric models (HOS) and dynamic human performance models (MicroSaint). Because of its ability to stimulate a task network with a computer, the MicroSaint portion of this tool allows for the prediction of human performance instead of merely a description of human performance normally associated with task analysis (Laughery, Plott, and Dahl, 1991). The HOS portion of the test provides an ideal environment to assess human performance parameters based on a detailed analysis of task components (Laughery, Plott, and Dahl, 1991). This tool may be extremely valuable for determining countermeasures and their impact on performance.

Unfortunately, a positive correlation exists between high mental workload, performance decrements, and error rates. As long duration space flights continue to increase in time and workload, it is reasonable to assume more stress will be placed on the crewmembers to accomplish assigned tasks. Consequently, it is also fair to assume that the frequency of human error will increase. One only needs to recall the recent mishaps on Mir where the crew suffered a series of problems which were attributable to human error. Two of the more serious and life-threatening errors include a misaligned docking attempt where a Progress capsule collided with one of the Mir capsules and the inadvertent removal of a critical computer power cable. As a direct result of these errors, the crewmembers had to struggle under less than optimal conditions to regain control of Mir. Although it can not be unequivocally determined that these errors were due to high workload and issues of prolonged isolation and confinement, these possible attributable causes can not be ruled out. Therefore, more research needs to be done to establish the relationship between high mental workload, isolation, and confinement on performance decrements and error rates.

Researchers at the Idaho National Engineering Laboratory (INEEL), in conjunction with Lockheed Martin Idaho Technologies, have developed measurement tools to identify the potential for human error. Furthermore, these tools are designed to evaluate and assess the effects of errors on system performance. Techniques will be developed to eliminate or reduce the error's impact on the system. Specifically, researchers have developed a framework and methodology called Framework Assessing Notorious Contributing

Influences for Error (FRANCIE) that facilitates the identification and modeling of human errors as well as the factors which contribute to the errors (Nelson, Haney, Ostrom, & Richards, 1995). In addition, researchers have developed a software tool called Tool for Human Error Analysis (THEA) which utilizes the FRANCIE framework and can be used by the aviation industry. In part, THEA was developed to "assist a user in exploring the potential impact that initial error events and failures to recover from those error events can have on an overall task performance" (Nelson et al., 1995). According to the authors, it is possible to apply this tool to space operations. Specifically, by encoding expert knowledge and experience regarding task performance and operations in a microgravity environment, it might be possible to assess the system, identify instances of human error, and develop new design specifications and task structures to reduce the potential for human error during the initial design phases.

Not only is mental workload a concern for LDSF, but physical workload must also be considered in relation to performance. According to Hayes et al. (1992), although the work related torques of EVA's and IVA's are relatively small, some repetitive tasks tend to be fatiguing. This repetitive work often produces local and systemic fatigue and soreness. Therefore, the limiting factor of crewmember performance during EVA may be based on the endurance and strength capacities of the hands, arms, and upper body. Thus, as the flight duration increases, effective completion of one and/or successive EVA mission tasks may be compromised.

In the industrial setting, many physical performance and workload measurement tools exist which are available for use by anyone with a need for a specific tool. Possible applications of these tools include biomechanical analyses, task analysis, research data acquisition (lifting and grip strength, response speed, hand performance, range of motion, etc.), and demonstrations of basic Human Factors principles such as Fitt's Law and Hick's Law. The following represents a list of some of the performance measurement tools available through the Human Performance Measurement, Inc., Arlington, Texas.

- 1) BEP I Central Processing and Upper Extremity Motor Control: measures more than 40 aspects of central processing and upper extremity performance including response speed, memory, ADL's, finger tapping speed, coordination, etc.
- 2) BEP II Lower Extremity Motor Control: lower extremity counterpart to the BEP I measuring lower extremity response speed to visual stimuli, tapping (movement) speed, leg movement speed, accuracy, and coordination (neuro-motor channel capacity).
- 3) BEP III Isometric Strength: includes a hand-held transducer for measuring strength of most functional units in the body using well-accepted procedures of manual muscle testing, and a compact device for grip strength measurement.

- 4) BEP IV Postural Stability: light-weight force platform system with embedded software
- 5) BEP V Steadiness/Tremor: a non-contacting capacitive 2-D sensor for upper extremity, lower extremity, and head steadiness.
- 6) BEP IX Tactile Sensation: a subsystem of instruments for measurement of tactile sensation, including a general controller/interface and up to five optional sub-units (vibration, thermal sense, touch/pressure, two-point discrimination, and electrical current perception.
- 7) BEP XI Speech/Hearing: acoustically based tests for a wide array of components in speech/hearing production systems
- 8) BEP IX Hand Performance: includes a rotary position and isometric strength transducer for measuring a wide variety of arm and hand performance resources, including twisting range of motion, strength, speed, finger/pinch strengths, and a joystick-type interface for measuring fine motor control performance in such tasks as tracking and rapid alternating movements

DESCRIPTION OF TOOLS

Within the scope of this report, a limited preliminary evaluation was conducted to identify the applicability of the existing tools for LDSF. The questions addressed were:

- 1) What is the format of use for each tool (i.e., paper/pencil or computer simulation)?
- 2) What are the applications of the tool (psychomotor performance, logic and decision making, mental workload, etc.)? and
- 3) Who developed the tool?

In addition, advantages and disadvantages of the tools as they apply to space missions will be included, if it can be determined. Note that the evaluation of physiological assessment tools is beyond the scope of this paper and therefore, the tools from Human Performance Assessment, Inc. will not be included in this section.

Habitability Tools

Mir debrief questionnaire of NASA crewmembers is a subjective questionnaire/interview used to assess habitability issues onboard Mir. This tool can be obtained through MOD-NASA-JSC and is a good reference to have after each flight.

Space Operations Issues Reporting Tool (SOIRT) is a form used to measure performance and habitability issues while in flight. This tool can be obtained through UTAF-FCSD-JSC and is also an important reference to have. Both of these tools can be evaluated and incorporated into future missions.

Performance Tools

PAWS (Performance Assessment Workstation), PATS (Psychophysical Assessment Test System) and UTC-PAB (Unified Tri-Service Cognitive Performance Assessment Battery) are tools to measure neuro-cognitive aspects of performance and require the use of computer simulation. PAWS can be obtained through NASA, PATS can be obtained through CSERIAC (Crew System Ergonomics Information Analysis Center), and the UTC-PAB can be obtained through DoD. These tools are quite useful for determining which task factors affect neuro/cognitive performance. Findings can be applied to the development of daily task schedules for crewmembers during LDSF to help reduce or eliminate performance decrements. However, in current form, these tools take a minimum of 1-2 hours to complete. Therefore, it is essential to allocate a dedicated time during the mission.

The AGARD STRES (Aerospace Research and Development battery of Standardized Tests for Research with Environmental Stressors) is developed as a computer simulation to measure psychomotor performance and performance related to logic and decision-making. This tool is developed through NATO. Findings can be applied to the development of daily task schedules for crewmembers during LDSF to help reduce or eliminate performance decrements. Similar to the previous tools, in current form, these tools take a minimum of 1-2 hours to complete so they should be included as activities within the mission timeline.

POMS (Profile of Mood States) is a subjective measurement technique and is developed by researchers at the University of Wisconsin to assess emotional state. It might be possible for crewmembers to complete the questionnaire in-flight to assess their emotional states at any given period.

Voice analysis is one tool that has received relatively minimal attention but may be very useful on LDSF. Voice analysis represents an extremely non-invasive measurement tool for the determination of stress among crewmembers. Because voice patterns can be analyzed any time the crew is in contact with ground controllers, no special measures need to be taken. Although not used by NASA, the technology for stress analysis is quite frequently used by the Russian ground controllers. An investigation into the Russian's use of this tool would be extremely beneficial.

SkyLab Medical Equipment Altitude Test (SMEAT) is a subjective questionnaire used to measure physical adaptation as it relates to performance and habitability issues. Although probably no longer applicable for future LDSF missions, this tool can be obtained through Experiment M487, NASA-JSC.

Workload Tools

There are a number of tools that have been developed over the past years to measure mental workload. A representative set of these tools will be discussed in the following paragraphs.

MINUTES (MINestoa Universal Task Evaluation System) is developed by researchers at the University of Minnesota to assess the effects of task load increments on performance. It uses a subjective evaluation of data and requires the crewmember to rate levels of perceived task load. Although slightly intrusive, this tool can be helpful in determining or predicting performance decrements during LDSF.

SWAT (Subjective Work Assessment Technique) is also created to test the effects of task load on performance. It can be obtained through CSERIAC. Like MINUTES, this tool is slightly intrusive but can be very helpful in determining or predicting performance decrements during LDSF.

MicroSaint is a computer simulation and can be used for task analysis purposes to create a dynamic human performance model. It can be obtained through Micro Analysis and Design, Inc. During ground-based simulations, this tool can be extremely useful in predicting the effects of task load on crewmembers over time. By studying the findings, it would be possible to redesign crew scheduling and/or experimental procedures to eliminate excessive task loads.

MicroSaint/HOS (Human Operator Simulator) is also developed by Micro Analysis and Design, Inc. and combines MicroSaint with an anthropometric evaluation of the person performing a given task. This combined tool adds the extra advantage of viewing crew body positioning and movement during task completion. The resultant posture can be evaluated from a human factors point of view, and if warranted, equipment can be redesigned to eliminate postural fatigue.

FRANCIE (Framework Assessing Notorious Contributing Influences for Error) is developed by Lockheed Martin Idaho Technologies and the Idaho National Engineering Laboratory (INEEL) as a computer framework that facilitates the identification and modeling of human error. They also developed a software tool, THEA (Tool for Human Error Analysis) which uses the FRANCIE framework to evaluate the potential impact of human error on a system. These tools can be extremely useful in identifying potential sources of error onboard space missions. Once identified, these errors can be designed out of the system to reduce the potential for life-threatening situations due to human error.

See Table 1 for further information regarding the tools.

Table 1. Habitability and Performance Measurement Tools

Factor to be Measured	Tool/Method of Data Collection	Mission/Task Outcome Affected	What/How Outcome Measured	Resource	Enabling Technology Needed	Use With Other Tools
Physical Adaptation	SMEAT	Performance	Subjective Evaluation	NASA	None	Yes
Habitability	MIR	Habitability/Performance	Subjective Evaluation	OC2/DO47: JSC	None	Yes
	Questionnaire/Debrief SOIRT	Habitability/Performance	Subjective Evaluation	FCSD/UTAF: JSC	None	Yes
Neuro/Cognitive	PAWS	Performance	Computer Simulation	NASA	Compatible Software System	N/A
	PATS	Performance	Computer Simulation	CSERIAC	Compatible Software System	N/A
	UTC-PAB	Performance	Computer Simulation	DoD	Compatible Software System	N/A
Psychomotor	AGARD STRES	Performance	Computer Simulation	NATO	Compatible Software System	N/A
Logic/Decision Making	ACARD STRES	Performance	Computer Simulation	NATO	Compatible Software System	N/A
Mood/Emotional State	POMS	Habitability/Performance	Subjective Evaluation	University of Wisconsin RSA	None	Yes
	Speech Analysis Techniques	Habitability/Performance	Physiological Evaluation		Proper Equipment	Yes
Mental Workload	MINUTES	Performance	Subjective Evaluation	University of Minnesota CSERIAC CSERIAC Dundee University MicroAnalysis & Design	None	Yes
	SWAT	Performance	Subjective Evaluation		None	Yes
	NASA-TLX	Performance	Subjective Evaluation		None	Yes
	DSSQ	Performance	Subjective Evaluation		None	Yes
	MicroSaint	Performance	Computer Simulation		Compatible Software System	N/A
	MicroSaint/HOS	Performance	Computer Simulation		Compatible Software System	N/A
	FRANCIE	Performance	Computer Simulation	L M Idaho & INEEL	Compatible Software System	N/A
	THEA	Performance	Computer Simulation	L M Idaho & INEEL	Compatible Software System	N/A

CONCLUSIONS

The area of habitability and performance has been extensively researched and studied. Despite the breadth of knowledge on the topic, little can be applied directly to LDSF. The majority of the ground-based performance tools require active participation on the part of the human subject and can be time consuming for use in-flight. Additionally, these tools are descriptive in nature, i.e., they are designed to define and assess which factors lead to performance decrements only after the fact. They can not be used to *predict* factors associated with performance problems. One notable exception is the MicroSaint/HOS tool developed by MicroAnalysis and Design, Inc. MicroSaint/HOS computer based simulation combines anthropometric measurements and task analysis to predict performance under various conditions. By simulating various conditions under which crewmembers will perform tasks, it is possible to determine which factors will cause performance decrements in-flight. In turn, these factors can be addressed prior to the actual LDSF. However, it is important to note that this tool is used for the prediction of *physical* performance and can not predict decrements in *mental* performance.

Because the other tools are not predictive in nature, it does not mean they are of no value for the space program. On the contrary, the information gathered on habitability and performance can be applied to future missions to ultimately improve conditions during LDSF. In fact, it is important to keep abreast of the research literature so new findings regarding habitability and performance can be applied as they become available.

Unfortunately, little information from the RSA and ESA was found in the way of mission reports, mission summaries, and crew debriefs. Without these reports, it is impossible to make cross-cultural determinations involving issues of habitability and performance. This information is critical for the International Space Station (ISS). In order to avoid the habitability and performance problems found on previous long-duration space flights, any information relating to these issues needs to be reviewed and applied when applicable.

More information was found regarding workload and performance than habitability. One possible explanation could be that there has not been enough research in the area of habitability, and therefore no conceptual models exist to quantitatively predict how various factors will affect habitability. Future work should investigate how various factors affect habitability and based on the findings, create a conceptual framework for predicting how habitability impacts performance and vice versa.

RECOMMENDATIONS

The implementation of habitability and performance assessment on LDSF can be viewed as a progression of three phases: planning, testing, and implementation (see Figure 2 for the proposed implementation process). The implementation process consists of the following critical steps:

- 1) Establish working knowledge of the current performance models/tools
- 2) Determine whether the tool needs to be developed (i.e., Does it exist and can it be refined for use in LDSF? If not, new tools should be developed for use during LDSF)
- 3) Test and validate the identified models and tools
- 4) Define and test potential countermeasures
- 5) Create flight rules to monitor continuously habitability and performance.

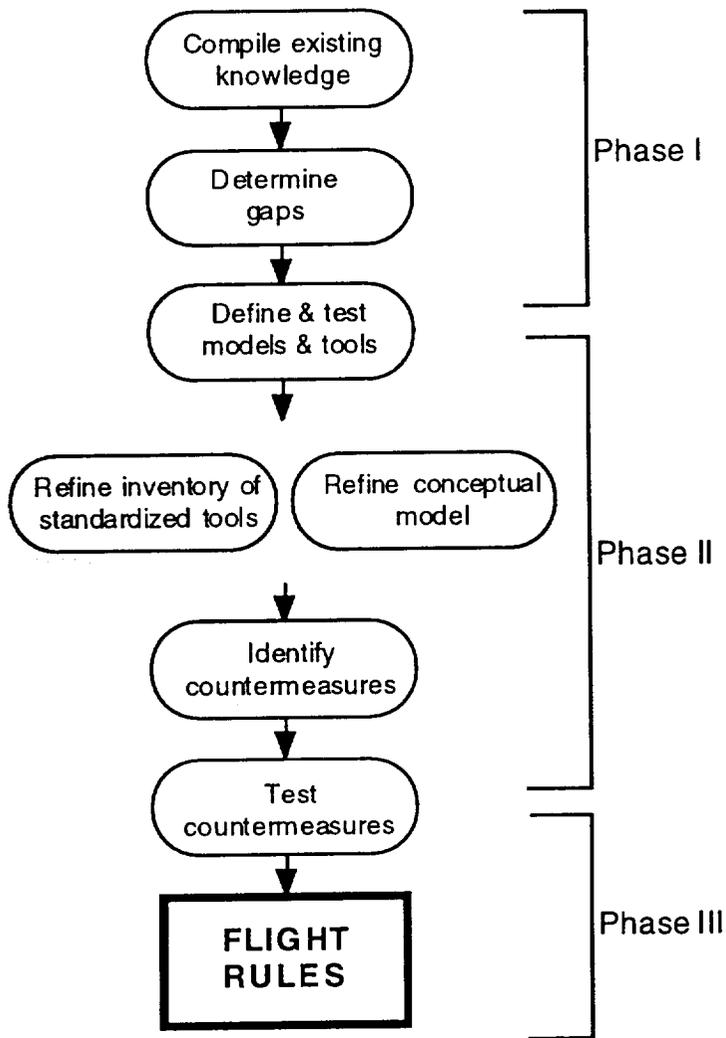


FIGURE 2. Implementation plan.

The primary purpose of the planing phase is to expand the current knowledge of habitability and performance factors which affect crewmembers during LDSF. More research needs to be done to determine the impact of various factors on performance and habitability. For example, how severe must the crew's lack of situational awareness be before there is an effect on performance? When do the first indications of decision making deficits appear during LDSF, and are there effective ways to counter these effects? Likewise, it is essential to obtain crew debriefs, mission debriefs, and other sources of flight information from the RSA, ESA, and other international space agencies. Lessons learned by international crew onboard Mir should be incorporated in future mission planning so the same problems experienced on Mir do not happen with other LDSF. Also, further contact needs to be established with the developers of existing tools to evaluate and assess the tool's applicability to space. Once received, all the information can be incorporated with existing NASA protocols to establish the best possible working and living environment aboard the ISS and future LDSF. Table 2 depicts a summary of significant activities for the planning phase.

Table 2. Significant activities for each phase.

<p><u>Planning Phase</u></p> <ul style="list-style-type: none"> • Acquire crewmember participation • Develop standardized questionnaires • Further investigate potential RSA & ESA sources • Research situational awareness <p><u>Testing Phase</u></p> <ul style="list-style-type: none"> • Gather and evaluate in-house tools • Investigate commercial tools • Investigate voice analysis techniques • Develop and test new tools <p><u>Implementation Phase</u></p> <ul style="list-style-type: none"> • Inventory standardized tools • Apply findings to future LDSF • Create prediction models & effective countermeasures • Build more flex-time into mission schedules

The second phase, or testing phase, should involve the acquisition and evaluation of the tools identified in the planning stage. These tools should be evaluated for their applicability to space, particularly LDSF. Once those tools are selected, ground-based testing of the tools can be conducted. This testing should encompass preliminary laboratory testing, evaluations using the flight simulations and analogs, and eventually Shuttle/Mir experiments such as Detailed Supplementary Objectives (DSO) and Risk Mitigation Experiments (RME). Likewise during this phase, potential countermeasures can be

developed and assessed for their impact on performance and habitability. See Table 2 for the phase II significant activities.

The third and final phase, implementation, should occur once testing is complete (see Table 2 for significant tasks). At this point, the standardized measurement tools should be used onboard every mission. Specifically, the tools and the appropriate performance countermeasures should be manifested as a mission task for the long-term research of the previously identified habitability and performance issues. It is important for these tools to have a simple interface since crewmembers will be using these tools on LDSF without constant contact with a primary investigator (PI). The crewmembers need to be able to pick up a tool and promptly learn or know how to use it. Also, the data collected from each crewmember should be easily retrievable so crewmembers can check their performance over time, and if needed, instigate countermeasures. When not being evaluated, the tool should provide the capability to compress the raw data for easy and efficient storage. A final goal of the implementation stage should include the creation of a conceptual model of habitability and performance to predict future problems during LDSF. Refer back to Figure 1 for some of the habitability issues which need to be investigated for their effect on performance. It is essential to determine which habitability factors have the most impact on performance and vice versa. Only then can an effective quantitative predictor model be established for use on LDSF.

Acknowledgments

The authors wish to thank Michele Segal for her help in contacting potential sources and reviewing some of the literature findings discussed in the paper. The authors also wish to thank Tina Holden, Cindy Chmielewski and Melissa Meingast for their help in proof-reading and editing.

References

Advisory Group for Aerospace Research and Development. (1989). Human performance assessment methods. AGARDograph no. 308, Neuilly-Sur-Seine.

Bittner, A. C., Carter, R. C., Kennedy, R. S., Harbeson, N. M., & Krause, M. (1986). Performance evaluation tests for environmental research (PETER): evaluation of 114 measures. Performance Motor Skills, 63, 683-708.

Blagrove, M., Alexander, C., & Horne, J. A. (1995). The effects of chronic sleep reduction on the performance of tasks sensitive to sleep deprivation. Applied Cognitive Psychology, 9, 21-40.

Boeing Habitability Report (28 February 1983). Space station habitability report.

Bonnet, M. H. & Arand, D. L. (1994). The use of prophylactic naps and caffeine to maintain performance during a continuous operation. Ergonomics, 37, 1009-1020.

Elsmore, T. F., Hegge, F. W., Naitoh, P., Kelly, T., Schlangen, K., & Gomez, S. (1995). A comparison of the effects of sleep deprivation on synthetic work performance and a conventional performance assessment battery. Naval Health Research Center. San Diego, CA.

Guidi, M. A. & Holland, D. (1992). Operational and Human Factors implications of physiological deconditioning in long duration space flight. Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting. Santa Monica: CA.

Hancock, P. A. & Williams, G. (1993). Effect of task load and task load increment on performance and workload. International Symposium on Aviation Psychology, 7th Annual Meeting. Columbus, Ohio.

Hart, S. G. & Staveland, L. E. (1988). Development of a multidimensional workload rating scale. Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload. Amsterdam: Elsevier.

Haslam, D. R. (1982). Sleep loss, recovery sleep, and military performance. Ergonomics, 25, 163-178.

Hayes, J. C., Stewart, D. F., Harris, B., Siconolfi, S., Greenisen, M., & LaRochelle, F. (1992). Program development for exercise countermeasures. Presented at the 22nd International Conference on Environmental Systems. Seattle Washington, July 13-16.

Holland, D. (1995). Consideration of group situation awareness and contextual elements for enhancing the human factor during long-duration space flight. AIAA Space Programs and Technologies Conference. Sept. 28-26, Huntsville, Alabama.

Human Performance Measurement, Inc., P.O. Box 1996, Arlington, Texas 76004-1996. Phone Number 817-640-7178: Fax 817-633-3142.

Kozlovskaya, I.B., Burlachkova, N.I., Ganchev, G., Gatev, P., Gerstenbrand, F., & Berger, M. (1993). Mechanisms of sensory-motor adaptation to weightlessness. Paper presented to the 10th IAA Man in Space Symposium, Tokyo, April 19-23.

Laughery, K. R., Plott, C., & Dahl, S. (1991). Integrating task networks and anthropometric models. Human-Centered Technology for Maintainability: Workshop Proceedings, Armstrong Laboratory.

McNair, D. M., Lorr, M., & Dropplemann, L. F. (1991). Profile of Mood States Manual. San Diego: Educational and Testing Service.

Manzey, D. (1989). Workload assessment by secondary tasks and the multi-dimensionality of human information processing resources. DLR-Institute for Aerospace Medicine, Department of Aviation and Space Psychology. Hamburg, Federal Republic of Germany.

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1993). Behavioral aspects of human adaptation to space: analyses of cognitive and psychomotor performance in space during an 8-day space mission. Clinical Investigator, 71, 725-731.

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1995). Dual task performance in space: Results from a single-case study during a short term space mission. Human Factors, 37, (4), 667-681.

NASA-JSC (1976). Skylab experiment M487: Habitability/Crew Quarters. NASA Technical Memorandum.

NASA-JSC (1976). Skylab experience bulletin No. 26: The methods and importance of man-machine engineering evaluations in zero-g. Nelson, W. R., Haney, L. N., Ostrom, L. T., & Richards, R. E. (199?). Structured methods for identifying and correcting potential human factors errors in space operations.

O'Donnell, R. D. & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), Handbook of perception and human performance. Vol. II. Cognitive processes and performance. New York: Wiley.

Ogden, G. D., Levine, J. M., & Eisen, E. J. (1979). Measurement of workload by secondary tasks. Human Factors, 21, 529-548.

Raglin, J. S. (1993). Overtraining and staleness: psychometric monitoring of endurance athletes. R. B. Singer, M. Murphey & L. K. Tennant (Eds.) Handbook of Research on Sport Psychology. New York: Macmillan.

Raglin, J. S. & Morgan, W. P. (1994). Development of a scale for use in monitoring training-induced distress in athletes. International Journal of Sports Medicine, 15, 84-88.

Salvendy, G. (1987). Handbook of Human Factors. New York: John Wiley & Sons.

Schlegel, R. E., Shehab, R. L., Schiflett, S. G., Eddy, D. R., & Gilliland, K. (1994). Astronaut baseline practice schedules for the NASA performance assessment workstation (PAWS). Aerospace Medical Association Abstracts, No. 166, 65th Annual Meeting, San Antonio, Texas, May 8-12.

Stuster, J. (1996). Bold endeavors: lessons from polar and space exploration. Annapolis: Naval Institute Press.

Veltman, J. A. & Gaillard, A. (1996). Physiological workload reactions to increasing levels of task difficulty. Report: TNO Human Factors Research Institute. 1-28.